

MKG-FENN: A Multimodal Knowledge Graph Fused End-to-End Neural Network for Accurate Drug–Drug Interaction Prediction

Di Wu^{1,2}, Wu Sun¹, Yi He^{3*}, Zhong Chen⁴, Xin Luo²

¹College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

²College of Computer and Information Science, Southwest University, Chongqing 400715, China

³Department of Computer Science, Old Dominion University, Norfolk, VA 23529, USA

⁴School of Computing, Southern Illinois University, Carbondale, IL 62901, USA

wudi.cigit@gmail.com, s220201081@stu.cqupt.edu.cn, yihe@cs.odu.edu, zhong.chen@siu.edu, luoxin21@gmail.com

Abstract

Taking incompatible multiple drugs together may cause adverse interactions and side effects on the body. Accurate prediction of *drug-drug interaction* (DDI) events is essential for avoiding this issue. Recently, various artificial intelligence-based approaches have been proposed for predicting DDI events. However, DDI events are associated with complex relationships and mechanisms among drugs, targets, enzymes, transporters, molecular structures, etc. Existing approaches either partially or loosely consider these relationships and mechanisms by a non-end-to-end learning framework, resulting in sub-optimal feature extractions and fusions for prediction. Different from them, this paper proposes a *Multimodal Knowledge Graph Fused End-to-end Neural Network* (MKG-FENN) that consists of two main parts: *multimodal knowledge graph* (MKG) and *fused end-to-end neural network* (FENN). First, MKG is constructed by comprehensively exploiting DDI events-associated relationships and mechanisms from four knowledge graphs of drugs-chemical entities, drug-substructures, drugs-drugs, and molecular structures. Correspondingly, a four channels graph neural network is designed to extract high-order and semantic features from MKG. Second, FENN designs a multi-layer perceptron to fuse the extracted features by end-to-end learning. With such designs, the feature extractions and fusions of DDI events are guaranteed to be comprehensive and optimal for prediction. Through extensive experiments on real drug datasets, we demonstrate that MKG-FENN exhibits high accuracy and significantly outperforms state-of-the-art models in predicting DDI events. The source code and supplementary file of this article are available on: <https://github.com/wudi1989/MKG-FENN>.

Introduction

The joint use of multiple drugs is very common in clinical care (Zitnik, Agrawal, and Leskovec 2018). While incompatible multiple drugs may cause adverse drug-drug interaction (DDI) events that have harmful side effects on the body (Vilar et al. 2014). For example, taking Abemaciclib alongside Bosutinib can lead to an increase in serum concentrations of Abemaciclib; Conversely, if Abemaciclib is taken at the same time as Clemastine, Abemaciclib’s metabolism

*Corresponding author: Dr. Yi He (yihe@cs.odu.edu)
Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

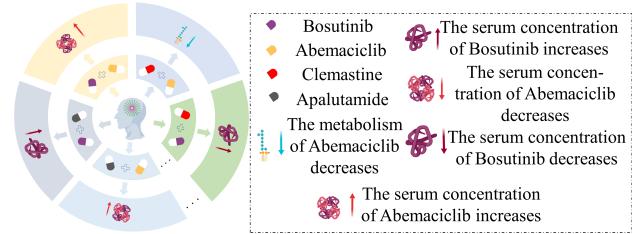


Figure 1: The examples of taking incompatible multiple drugs together may cause adverse interactions and side effects on the body.

may be impaired, as illustrated in Figure 1. To avoid the negative consequences associated with DDI events and maximize their benefits, it is crucial to proactively identify such interactions in advance.

DDI events prediction is essential for addressing this issue. Recently, various artificial intelligence-based approaches have been proposed for achieving DDI events prediction, such as utilizing graph neural networks (GNNs) to analyze the structural similarity of chemicals (Huang et al. 2020), employing semi-supervised learning to extract valuable information from labeled and unlabeled drug data (Chu et al. 2019), and leveraging knowledge graph for predicting the pharmacological effects of multi-typed DDIs (Yu et al. 2021). Notably, DDI events are associated with complex relationships and mechanisms among drugs, targets, enzymes, transporters, molecular structures, etc. (Cui et al. 2020; Lin et al. 2020). To exploit such complex relationships and mechanisms, some efforts have been made (Deng et al. 2020), including drug features analyzed (Chu et al. 2019), graph learning-based (Yu et al. 2021), and hybrid modeling (Lyu et al. 2021) approaches. However, these approaches either partially or loosely consider such relationships and mechanisms by a non-end-to-end learning framework, making the cross-modality complementarity and potential relationships between different aspects of DDI events be not well exploited. As a result, the feature extractions and fusions of DDI events are sub-optimal for predictions.

To address this issue, this study proposes a novel *Multi-*

modal Knowledge Graph Fused End-to-end Neural Network (MKG-FENN) for accurate DDI events prediction. MKG-FENN has two main parts: multimodal knowledge graph (MKG) and fused end-to-end neural network (FENN). First, MKG is to comprehensively exploit DDI events associated relationships and mechanisms by constructing four knowledge graphs of drugs-chemical entities, drugs-substructures, drugs-drugs, and molecular structures, where a four channels GNN is designed to extract high-order and semantic features from MKG. Second, FENN designs a multi-layer perceptron to fuse the extracted features by an end-to-end learning framework. With such designs, MKG-FENN can guarantee that the feature extractions and fusions of DDI events are comprehensive and optimal for achieving accurate predictions.

Specific contributions of this paper are as follows:

- i) This is the first study to comprehensively exploit various relationships and mechanisms of drugs-chemical entities, drugs-substructures, drugs-drugs, and molecular structures to predict DDI events in end-to-end learning way.
- ii) A novel and highly-accurate MKG-FENN model is proposed and elaborated for DDI events prediction.
- iii) Extensive experiments on real drug datasets are conducted to evidence the viability, effectiveness, and superiority of the proposed MKG-FENN model.

Due to page limit, some contents including *notations, algorithm designs, details of rival models and implementation, and certain experimental results* are deferred to the **Supplementary File**, available as an electronic companion.

Related Work

Drug Features Analyzed Methods

Drug features analysis is a difficult but important process for DDI events prediction. Some studies assume that similar drugs may exhibit similar DDIs and propose to learn accurate and interpretable similarity measurements from different types of drug features for DDI prediction (Deng et al. 2020). DeepDDI (Ryu, Kim, and Lee 2018) is a deep learning method to learn drug pairs and drug-food constituent pairs for DDIs prediction. MDF-SA-DDI (Lin et al. 2022b) introduces a DDI event prediction model based on multi-source drug fusion, multi-source feature fusion, and transformer self-attention for offline drug feature learning. ML-RDA (Chu et al. 2019) is developed to effectively utilize multiple features of drugs by incorporating a novel unsupervised disentangling loss called CuXCov. DeSIDE-DDI (Kim and Nam 2022) is a deep learning-based framework that focuses on interpreting the underlying genes in DDIs analysis. Recently, a multi-type DDI prediction model named MDDI-SCL (Lin et al. 2022a) is presented by supervised contrastive learning and three-level loss functions. Nonetheless, most of the drug features analyzed methods focus on acquiring the extensive attributes and features of drugs while neglecting the topological information and semantic relationships among drugs, targets, enzymes, transporters, molecular structures, etc. In comparison, the proposed MKG-FENN model can extract such high-order topological information

and semantic relationships, which is of great help to DDI events prediction.

Graph Learning-Based Methods

Graph embedding-based. Currently, there are many graph embedding methods used for effective network-based features in DDI prediction. These methods fall into three categories. The first category involves models that use the adjacency matrix as input to learn latent embeddings through matrix decomposition (Shi et al. 2019).

The second category focuses on generating sequences of nodes through random walks and learning node representations based on these sequences (Ribeiro, Saverese, and Figueiredo 2017). The final category utilizes diverse neural architectures and graph data as input to capture higher-order connectivity patterns and leverage rich drug network information (Tang et al. 2015; Wang, Cui, and Zhu 2016).

Knowledge graph-based. The utilization of knowledge graphs has significantly advanced research in various domains, including relation inference and recommendation (Wang et al. 2019). KGNN (Lin et al. 2020) successfully integrated graph convolutional networks with neighborhood sampling to effectively extract neighborhood relations. SumGNN (Yu et al. 2021) introduced a graph summarization module for subgraphs to extract manageable pathways. LaGAT (Hong et al. 2022) proposed a link-aware graph attention method that generates multiple attention pathways for drug entities based on different links between drug pairs. DDKG (Su et al. 2022) extended this idea by learning drug embeddings from their attributes in the KG and considering neighboring node embeddings and triple facts simultaneously using an attention mechanism.

Molecular graph-based. This type method encompasses predicting molecular properties (Wang et al. 2022) and molecular interactions (Li et al. 2022). MFFGNN (He et al. 2022) combines the topological structure within molecular graphs with the interaction relationship between drugs and the local chemical context in SMILES sequences. Furthermore, Molormer (Zhang et al. 2022) leverages the two-dimensional structures of drugs as input and encodes the molecular graph with spatial information using a lightweight attention mechanism.

Note that although these graph learning-based methods have delved into the higher-order structure and semantic relationships of drugs, they still partially consider these relationships. In comparison, the proposed MKG-FENN model has comprehensively exploited various relationships and mechanisms from drugs, chemical entities, and molecular structures.

Hybrid Modeling Methods

Hybrid modeling is more effective than individual models (Chen et al. 2021). MDNN (Lyu et al. 2021) design a two-pathway framework including a drug knowledge graph pathway and a heterogeneous features pathway to for predicting DDI events. Deepika and Geetha (Deepika and Geetha 2018) employed a semi-supervised learning framework that incorporated network representation learning and

meta-learning techniques on different drug datasets. Chen et al. (Chen et al. 2021) introduced a multi-scale feature fusion deep learning model called MUFFIN to learn drug representations from both the drug structure and a biomedical knowledge graph. However, these hybrid modeling methods are non-end-to-end learning frameworks, resulting in sub-optimal feature extractions and fusions for DDI events prediction. In comparison, the proposed MKG-FENN has an end-to-end learning way to fuse the extracted features, which guarantees the feature extractions and fusions of DDI events always to be comprehensive and optimal.

Preliminaries

DDI Matrix. The DDI matrix represents drug-drug interaction events and is denoted as $\mathcal{Y} \in (0, y_{ij})^{N_d \times N_d}$, where N_d represents the number of drugs in the matrix. The label matrix consists of labels $y_{ij} \in \mathcal{L}$, where $\mathcal{L} = \{y_1, y_2, \dots, y_{N_l}\}$ denotes the set of possible labels, and N_l represents the number of event types. Each element $y_{ij} \in \mathcal{L}$ in the DDI matrix indicates whether there is a presence or absence of an interaction event between drug d_i and drug d_j . If $y_{ij} = 0$, it signifies that there is no interaction event between drug d_i and drug d_j . Please note that in our experiment, there are a total of 65 classes of DDI events, such as "serum concentration increase" and "fluid increase", and we have assigned a unique identifier to each event class.

Drug Knowledge Graph. The drug knowledge graph is a specialized knowledge graph for predicting DDI events. It is represented by the tuple $\mathcal{G} = (\mathcal{D}, \mathcal{R}, \mathcal{T})$. In this representation, \mathcal{D} denotes a subset of drug entities, \mathcal{T} represents a subset of tail entities, which are related to drugs (such as targets), and \mathcal{R} denotes the set of relations between drugs and tail entities. The drug knowledge graph is defined as a set of tuples (d, r_{dt}, t) , where each tuple represents a connection between a drug entity d , a relation r_{dt} , and a tail entity t . These connections exist if and only if the drug entity is in the set \mathcal{D} , the relation is in the set \mathcal{R} , and the tail entity is in the set \mathcal{T} . By constructing and analyzing the drug knowledge graph, insights can be drawn regarding the relationships between drugs and their corresponding tail entities.

DDI Events Prediction. Our objective is to predict specific interaction events between drug d_i and drug d_j using the DDI events matrix \mathcal{Y} and drug knowledge graph \mathcal{G}^i , where i represents the number of drug knowledge graphs. To achieve this, we learn a prediction function denoted as $\hat{y}_{ij} = \Gamma(d_i, d_j | \Theta, \mathcal{Y}, \mathcal{G}^i)$, where \hat{y}_{ij} represents the probability of an event occurring between drug d_i and drug d_j . The function Γ incorporates the model parameters Θ and the information from \mathcal{Y} and \mathcal{G}^i to make accurate predictions.

Proposed Method

Overview. The architecture of MKG-FENN is depicted in Figure 2, which consists of two main functional parts: multimodal knowledge graph (MKG) and fused end-to-end neural network (FENN). MKG employs a GNN to extract the features of both topological structure and semantic relationships from the various knowledge graphs associated with

DDI events. FENN aims to effectively exploit the potential complementarity and relationships among the extracted features in an end-to-end learning way.

Multimodal Knowledge Graph

Constructing Drug Knowledge Graphs. We construct four knowledge graphs of drugs-chemical entities, drug-substructures, drugs-drugs, and molecular structures. Each drug knowledge graph is represented in the form of triples, denoted as $\langle \text{drugs}, \text{relationships}, \text{entities} \rangle$. These triples capture the relationships between drugs and various entities in the knowledge graph. We used Unified Medical Language System and DrugBank ID for unified identifier system and knowledge graph construction.

First, the data of drug knowledge graphs are sourced from DrugBank. We collect drug-related information such as transporters and targets to serve as the entities in this part. To establish the relationships between drugs and entities, we assign the general function of the entity as the corresponding relationship. For instance, let's consider the drug Lovastatin. If there is a transporter called Serum albumin, and its general function is Toxic substance binding, we would create the following triplet: $\langle \text{Lovastatin}, \text{Toxic substance binding}, \text{Serum albumin} \rangle$.

Second, the data of drug knowledge graphs are sourced from the SMILES attribute in DDIMDL (Deng et al. 2020). In this process, the SMILES attribute of drugs is considered as entities, while the relationship between drugs and entities is represented by "including".

Third, the data of drug knowledge graphs are derived from DDI events matrix. This dataset is known for its substantial size and rich information. In this data, we can gather information about the other drug that each drug can interact with. We treat these drugs as entities, and the specific interaction event as the corresponding relationship.

Finally, the drug knowledge graphs are constructed based on the Molecular ACCess System (MACCS) bonds along with 13 MACCS bonds and 7 other molecular features (Baranwal et al. 2020). These MACCS bonds and molecular features are treated as entities of the drug, where the values indicating their belonging frequencies are denoted as relationships. For instance, Glucosamine possesses three molecular substructures of NumSaturatedRings, yielding $\langle \text{Glucosamine}, 3, \text{NumSaturatedRings} \rangle$.

Extracting Features by the GNN Layer. The purpose of using the GNN layer is to obtain the topological structure and semantic relationships of drugs. In this article, the drug knowledge graph is transformed into a matrix representation. The initial representation matrix of the drug knowledge graph, denoted as \mathcal{G}^i , is as follows:

$$E^i_{\mathcal{G}^i} = [\underbrace{e_{d_1}^{(0)}, \dots, e_{N_d}^{(0)}}_{\text{drug-embedding}}, \underbrace{e_{r_1}^{(0)}, \dots, e_{N_r}^{(0)}}_{\text{relation-embedding}}, \underbrace{e_{t_1}^{(0)}, \dots, e_{N_k}^{(0)}}_{\text{tail-embedding}}] \quad (1)$$

In the formula, $E^i_{\mathcal{G}^i}$ represents the i -th knowledge graph's initial representation matrix, where i ranges from 1 to 4. N_d , N_r and N_k indicate the number of drugs, relationships, and tail entities, respectively. $e_d^{(0)} \in R^d$,

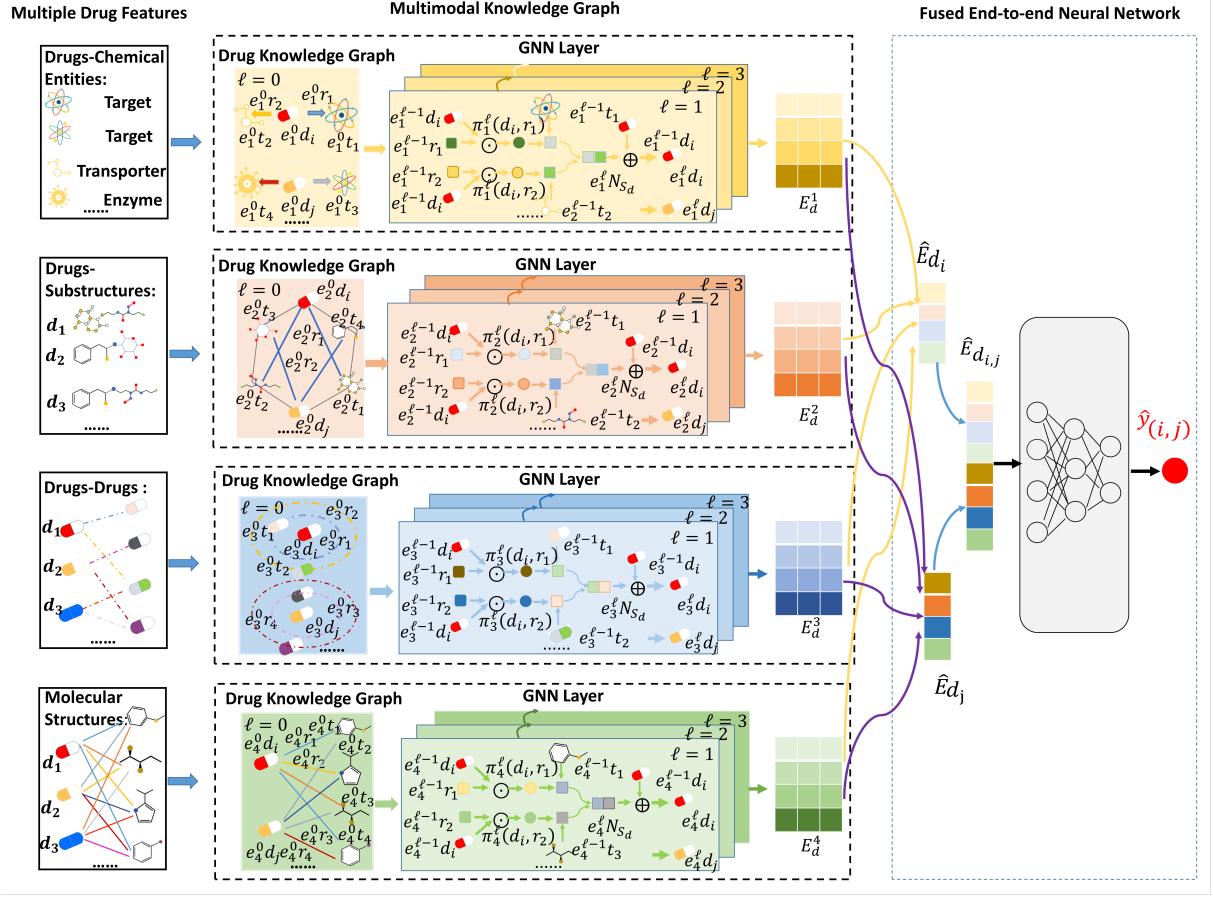


Figure 2: The overall structure of the proposed MKG-FENN model.

$e_r^{(0)} \in R^d$ and $e_t^{(0)} \in R^d$ represent the initial embeddings for drugs, relationships, and tail entities, where d is the embedding dimension of the drug knowledge graph.

For each drug d_i , a fixed-size neighborhood of samples is uniformly sampled as the drug's neighbors instead of considering all tail entities. The fixed-size neighborhoods are represented by $N_s(d_i)$.

Assuming the drug knowledge graph \mathcal{G} of drug d_i , we represent it in triples (d_i, r_{in}, t_n) , where t_n represents the neighborhood of drug d_i and r_{in} represents the semantic relationship within the neighborhood. To incorporate the semantics of relationships into drug representation learning, we calculate the semantic feature score between drug d_i and its corresponding neighborhood tail entity t_n using the following formula:

$$\pi_{(d_i, r_{in})}^{(l)} = \text{sum} \left[\left(e_{d_i}^{(l-1)} \odot e_{r_{in}}^{(l-1)} \right) W_1^{(p)} + b_1^{(p)} \right] \quad (2)$$

In the formula, $e_{r_{in}}^{(l-1)}$ represents the embedding of the relationship between drug d_i and tail entity t_n in the $(l-1)^{th}$ layer of the GNN. $e_{d_i}^{(l-1)}$ represents the embedding of drug d_i in the $(l-1)^{th}$ layer of the GNN. $W_1^{(p)}$ is the trainable

weight matrix, $b_1^{(p)}$ is the bias vector, and p represents the number of fully connected layers. The symbol \odot represents element-wise multiplication.

Next, we aggregate the embeddings of the neighborhood $N_s(d_i)$ by combining the embeddings of the neighborhood and the semantic feature scores. The aggregation function is defined as follows:

$$e_{N_s(d_i)}^{(l)} = \sum_{t_n \in N_s(d_i)} \pi_{(d_i, r_{in})}^{(l)} e_{t_n}^{(l-1)} \quad (3)$$

In the formula, $e_{t_n}^{(l-1)}$ represents the neighborhood embedding of drug d_i in the $(l-1)^{th}$ layer of the GNN. $\pi_{(d_i, r_{in})}^{(l)}$ represents the semantic feature score of drug d_i and the relationship in the $(l)^{th}$ layer.

The final step involves the aggregation process. To fuse the embedding of drug d_i with its corresponding neighborhood representation into a vector, we use the following fusion equation:

$$E_{d_i} = e_{d_i}^{(l)} = \sigma \left(\left(e_{d_i}^{(l-1)} \oplus e_{N_s(d_i)}^{(l)} \right) W_2 + b_2 \right) \quad (4)$$

where σ represents the ReLU activation function, and \oplus denotes the concatenation operation. $W_2 \in R^{(2d)*d}$ is the trainable weight matrix, and b_2 is a bias vector.

Note that the constructed four drug knowledge graphs share a consistent format. Similarly, we can apply the same methodology to calculate the representation of drug d_j using its corresponding knowledge graphs. By applying the formula and generating the drug representation, we can capture the relevant information and features of drug d_j within the drug knowledge graph.

Fused End-to-End Neural Network

To maximize the information from the four paths, we take into account their complementarity and correlation in the fused end-to-end neural network layer. For drug d_i , we can obtain drug representations corresponding to the four parts of the dataset in the graph neural network, represented by $E^1_{d_i}$, $E^2_{d_i}$, $E^3_{d_i}$ and $E^4_{d_i}$, respectively. The fusion of the four-part drug representations can be described by the following formula:

$$\hat{E}_{d_i} = E^1_{d_i} \oplus E^2_{d_i} \oplus E^3_{d_i} \oplus E^4_{d_i} \quad (5)$$

By following this approach, the final representations of the drug, denoted as \hat{E}_{d_i} , incorporate semantic relationships and topological information from the four parts of the dataset. Similarly, for drug d_j , we can utilize the same method to obtain its fused representations, denoted as \hat{E}_{d_j} .

Next, a multi-layer perceptron (MLP) is utilized to predict DDI events between different drugs. The formula for the MLP prediction is described as follows:

$$\hat{y}_{ij} = \sigma \left(\left(\hat{E}_{d_i} \oplus \hat{E}_{d_j} \right) W_3^{(q)} + b_3^{(q)} \right) \quad (6)$$

In the formula, $W_3^{(q)}$ represents the trainable weight matrix, $b_3^{(q)}$ denotes the bias vector, and q represents the total number of fully connected layers in the model. The activation function σ used in these layers is ReLU, which applies element-wise rectification and introduces non-linearity to the network.

To optimize the model, we incorporate a batch normalization layer to expedite convergence. Additionally, we introduce a dropout layer to mitigate overfitting and enhance the model's generalization capabilities. Furthermore, we apply ℓ_2 regularization to counteract overfitting tendencies. For the optimization process, we employ the Adam optimizer and utilize cross-entropy as the loss function.

Experiments

In the subsequent experiments, we conduct three kinds of experiments as follows: predicting DDI events between known drugs (Task 1), between known drugs and new drugs (Task 2), and among new drugs (Task 3). Then, we aim to answer the following research questions (RQs):

- RQ.1. Can the proposed MKG-FENN model outperform state-of-the-art models in Task 1?
- RQ.2. How does the proposed MKG-FENN model compare to state-of-the-art models in Tasks 2 and 3?
- RQ.3. How do different channels impact the performance of the MKG-FENN model (ablation study)?
- RQ.4. How do different hyper-parameter settings affect the performance of the MKG-FENN model?

General Settings

Datasets. To validate the effectiveness of MKG-FENN, we collect real drug datasets that have four parts. **Part 1:** We obtained this portion from DrugBank¹ (version 5.1.7) based on DDI events. **Part 2:** This section was obtained from DDIMDL², and we converted the drug's smile characteristics into a knowledge graph. **Part 3:** This section is derived from DDIMDL, where we converted the DDI matrix into a knowledge graph. **Part 4:** Part 4 represents the MACCS (Molecular ACCess System) keys. We selected 13 MACCS keys and 7 other molecular features based on the work of (Baranwal et al. 2020). The four parts of datasets are employed to construct four drug knowledge graphs described in Figure 2. The detailed information of the datasets is provided in Table 1.

Dataset	Drug number	Entity number	Relationship number	Triple number
Part 1	572	825	235	6541
Part 2	572	583	1	70350
Part 3	572	572	65	74528
Part 4	572	20	13	11440

Table 1: The details of the dataset.

Evaluation Metrics. For model evaluation metrics, we employ multi-class classification evaluation metrics, including accuracy (ACC), area under the precision-recall curve (AUPR), area under the ROC curve (AUC), F1 score, precision (Pre), and recall (Rec) (Lyu et al. 2021).

Baselines. The proposed MKG-FENN model is compared with six state-of-the-art related models: MDDI-SCL (Lin et al. 2022a), MDF-SA-DDI (Lin et al. 2022b), DDIMDL (Deng et al. 2020), MDNN (Lyu et al. 2021), Lee et al.'s methods (Lee, Park, and Ahn 2019), and Deep-DDI (Ryu, Kim, and Lee 2018). Additionally, several traditional classification methods are also considered, namely DNN, RF, KNN, LR (Deng et al. 2020), and GNN (Kipf and Welling 2016).

Hyper-Parameter. In the experiment, the training was configured with the following settings: 120 epochs of iterations, a learning rate of 0.01, a batch size of 1024, an embedding size of 128, a N_s (neighborhood size) of 6 , and a

¹<https://go.drugbank.com/>

²<https://github.com/YifanDengWHU/DDIMDL>

Metric	MDDI-SCL	MDF-SA-DDI	DDIMDL	MDNN	Lee et al.' methods	DeepDDI	DNN	RF	KNN	LR	GNN	MKG-FENN	
Task 1	ACC	0.9378	0.9301	0.8852	0.9175	0.9094	0.8371	0.8797	0.7775	0.7214	0.7920	0.9142	0.9409
	AUPR	0.9782	0.9737	0.9208	0.9668	0.9562	0.8899	0.9134	0.8349	0.7716	0.8400	0.9691	0.9786
	AUC	0.9983	0.9989	0.9976	0.9984	0.9961	0.9961	0.9963	0.9956	0.9813	0.9960	0.9989	0.9989
	F1	0.8755	0.8878	0.7585	0.8301	0.8391	0.6848	0.7223	0.5936	0.4831	0.5948	0.8332	0.8958
	Pre	0.8804	0.9085	0.8471	0.8622	0.8509	0.7275	0.8047	0.7893	0.7174	0.7437	0.8941	0.9132
	Rec	0.8767	0.8760	0.7182	0.8202	0.8339	0.6611	0.7027	0.5161	0.4081	0.5236	0.8012	0.8876
Win/Tie/Loss 6/0/0		5/1/0	6/0/0	6/0/0	6/0/0	6/0/0	6/0/0	6/0/0	6/0/0	6/0/0	5/1/0	64/2/0*	
Statistic p-value		0.0156	0.0313	0.0156	0.0156	0.0156	0.0156	0.0156	0.0156	0.0156	0.0313	-	
F-rank		10	10.5	6.17	8.17	7.25	3.75	5.17	2.33	1	3	8.83	11.83

Table 2: The comparison between MKG-FENN and its competitors in task 1, including the Win/Tie/Loss counts, Wilcoxon signed-ranks test, and Friedman test.

classification loss weight of $1e-08$. Additionally, the parameters were set to $l = 1$, $p = 2$, and $q = 3$. It is worth noting that l denotes the number of hidden layers in GNN. Our empirical study suggested using 1 layer, aligning with the recent issue of over-smoothing in GNN (Lyu et al. 2021).

Comparison Based on Known Drugs (RQ.1)

Task 1 holds significant importance for DDI prediction. Table 2 illustrates the comparison of our model with all the baselines in terms of the evaluation metric for Task 1. To gain deeper insights into these results, we conducted statistical analyses involving win/tie/loss analysis, the Wilcoxon signed-ranks test, and the Friedman test.

In Table 2, we can clearly see that our model outperforms all the other models across all evaluation metrics. This is evident from the total win/tie/loss cases, where our model has achieved a remarkable 64 wins and 2 tie, with no losses. Furthermore, all the p-values calculated for the comparisons are smaller than 0.05, indicating that the performance improvement of our MKG-FENN model is statistically significant compared to the other models, with a significance level of 0.05. Additionally, our model boasts the highest F-rank value, further solidifying its superiority.

In addition, we conducted a detailed analysis of MKG-FENN’s performance for each event in task 1 and calculated a metric score for each event using predicted scores and real labels. The AUPR scores and AUC scores for all predictive models are presented in Figure 3. The results depicted in Figure 3 clearly demonstrate that MKG-FENN consistently achieved higher AUPR and AUC scores compared to other models across the majority of events. To further scrutinize the comparison, we utilized boxplots to showcase the superior performance of our model in these events. Figure 4 illustrates the distribution of model performances. It is worth noting that our proposed model can achieve better performance than other models for small-sample data. For instance, we can accurately predict event number 62, which is not achievable by other comparative models.

Comparison Based on New Drugs (RQ.2)

Table 3 and Table 4 present the performance comparison between our model and baselines in tasks 2 and 3. For task 1, we employed five-fold cross-validation to divide the DDI

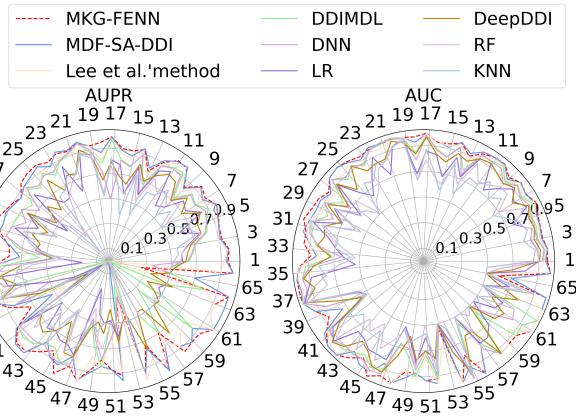


Figure 3: The AUPR scores and AUC scores of all prediction models for each event.

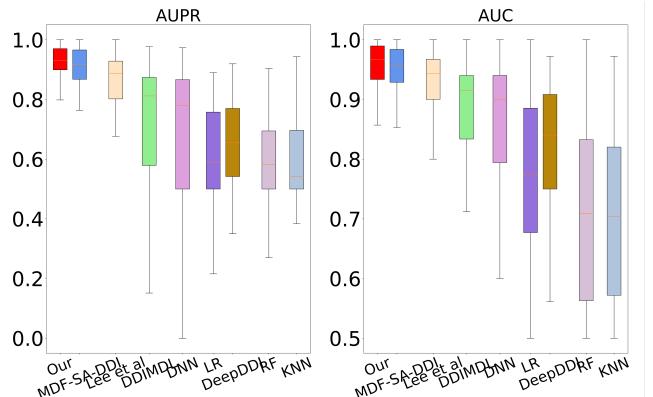


Figure 4: Boxplots displaying the AUPR and AUC of compared methods of each event.

event dataset into five subsets, with four subsets used for training and one subset for testing. In tasks 2 and 3, we initially divided the drug types into five parts, with one part consisting of new drugs. Subsequently, we further divided the data of new drugs in the DDI dataset to create a test set.

Based on the experimental results, it is evident that MKG-FENN outperforms the comparison models in most cases and achieves a higher F-rank value. Specifically, in tasks 2 and 3, the overall outcomes can be summarized as 31 wins, 0 ties, and 5 losses in favor of MKG-FENN. Moreover, the F-rank value of our model is 6.17, which achieves the highest F-rank value among all the models. These results demonstrate the superior performance of MKG-FENN.

Metric	MDDI-SCL	MDF-SA-DDI	DDIM-DL	Lee et al.'s methods	Deep DDI	DNN	MKG-FENN
ACC	0.6767	0.6633	0.6415	0.6405	0.5774	0.6239	0.6805
AUPR	0.6947	0.6776	0.6558	0.6244	0.5594	0.6361	0.7049
AUC	0.9634	0.9497	0.9799	0.9247	0.9575	0.9796	0.9673
F1	0.5304	0.5584	0.4460	0.5039	0.3416	0.2997	0.5394
Pre	0.6254	0.6547	0.5607	0.5388	0.3630	0.4237	0.6063
Rec	0.4814	0.5078	0.4319	0.4891	0.3890	0.2840	0.5106
Win/Tie /Loss	5/0/1	4/0/2	5/0/1	6/0/0	6/0/0	5/0/1	31/0/5*
p-value	0.0781	0.1563	0.0469	0.0156	0.0156	0.0313	-
F-rank	5.17	5.33	4.17	3	1.67	2.50	6.17

Table 3: The comparison results between MKG-FENN and its competitors in task 2.

Ablation Study (RQ.3)

To examine the influence of different channels, we conducted an ablation study by combining the different knowledge graphs of MKG-FENN. The results are presented in Table 5. Analyzing these results, it is evident that constructing a drug representation based on drug topology is a viable approach. We achieved satisfactory performance using a single-channel model. However, when employing a multi-channel fusion model, we observed a continuous performance improvement, ultimately leading to the best results with the MKG-FENN model. Including different channels aims to explore the representation of different aspects of drugs, thereby enriching the learning process regarding

Metric	MDDI-SCL	MDF-SA-DDI	DDIM-DL	Lee et al.'s methods	Deep DDI	DNN	MKG-FENN
ACC	0.4589	0.4338	0.4075	0.4097	0.3602	0.4087	0.4552
AUPR	0.3938	0.3873	0.3635	0.3184	0.2781	0.3776	0.4162
AUC	0.9053	0.8630	0.9512	0.8302	0.9059	0.9550	0.9149
F1	0.1919	0.2329	0.1590	0.2022	0.1373	0.1152	0.2186
Pre	0.2585	0.2715	0.2408	0.2216	0.1586	0.1836	0.2754
Rec	0.1678	0.2226	0.1452	0.2027	0.1450	0.1093	0.2131
Win/Tie /Loss	5/0/1	4/0/2	5/0/1	6/0/0	6/0/0	5/0/1	31/0/5*
p-value	0.0313	0.1563	0.0313	0.0156	0.0156	0.0469	-
F-rank	4.83	5.33	3.50	3.33	1.83	3	6.17

Table 4: The comparison results between MKG-FENN and its competitors in task 3.

	ACC	AUPR	AUC	F1	Pre	Rec	F-rank
P1	0.9172	0.9696	0.9988	0.8256	0.9011	0.7858	5.17
P2	0.9169	0.9693	0.9989	0.8165	0.8842	0.7846	4.08
P3	0.9169	0.9688	0.9989	0.8200	0.8933	0.7802	3.92
P4	0.9157	0.9680	0.9987	0.8230	0.9127	0.7825	4.17
P1+P2	0.9341	0.9758	0.9987	0.8707	0.9117	0.8492	10.08
P1+P3	0.9307	0.9742	0.9987	0.8666	0.9113	0.8436	6.83
P1+P4	0.9322	0.9748	0.9987	0.8669	0.9075	0.8465	7.67
P2+P3	0.9318	0.9749	0.9987	0.8689	0.9037	0.8468	7.67
P2+P4	0.9303	0.9743	0.9987	0.8661	0.9088	0.8416	6.33
P3+P4	0.9297	0.9741	0.9987	0.8700	0.9073	0.8457	6.50
P1+P2+P3	0.9344	0.9751	0.9986	0.8858	0.9199	0.8626	10.50
P1+P2+P4	0.9355	0.9755	0.9985	0.8847	0.9218	0.8648	11.17
P1+P3+P4	0.9349	0.9758	0.9985	0.8868	0.9161	0.8692	11.25
P2+P3+P4	0.9351	0.9757	0.9985	0.8765	0.9210	0.8542	10.50
P1+P2+P3+P4	0.9409	0.9786	0.9989	0.8958	0.9132	0.8876	14.17*

* P1, P2, P3, and P4 represent 4 parts of the dataset, respectively.

Table 5: The ablation study of MKG-FENN.

different drug topologies. Therefore, the fusion of multi-channel drug embeddings can enhance the prediction accuracy of DDI events.

Parameter Sensitivity Analysis (RQ.4)

In this study, we identified three crucial parameters: the size of the sampling neighborhood \mathcal{N}_s , the dimension of the drug embedding d in the drug knowledge graph, and the classification loss weight (CLW). To investigate the impact of these parameters, we conducted experiments while keeping the other parameters fixed. Finally, the model achieved optimal performance when $\mathcal{N}_s = 6$; the model performs best when $d = 128$; and the optimal CLW value for achieving the best model performance is $1e - 8$. Please refer to the **Supplementary File** to see details.

Conclusion

This paper proposes a novel MKG-FENN model for predicting DDI events. MKG-FENN utilizes a multi-channel GNN to effectively leverage both topological information and semantic relationships from four knowledge graphs of drugs-chemical entities, drugs-substructures, drugs-drugs, and molecular structures. Additionally, the MKG-FENN is designed as an end-to-end and tightly integrated architecture that learns to effectively integrate and converge information from the four constructed drug knowledge graphs. To evaluate the proposed MKG-FENN, extensive experiments are conducted on real drug datasets. The results demonstrate that the MKG-FENN significantly outperforms both traditional approaches and state-of-the-art models in predicting DDI events. In the future, we plan to improve MKG-FENN by incorporating some self-attention mechanisms.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 62176070, 62371438, and 62272078, in part by the Science and Technology Foundation of State Grid Corporation of China under grant 1400-202357341A-1-1-ZN, in part by the Chongqing Technical Innovation and Application Development Special Project under grant CSTB2023TIAD-KPX0037, in part by the Commonwealth Cyber Initiative (CCI), and in part by the National Science Foundation (NSF) under Grants IIS-2245946 and IIS-2236578.

References

- Baranwal, M.; Magner, A.; Elvati, P.; Saldinger, J.; Violi, A.; and Hero, A. O. 2020. A deep learning architecture for metabolic pathway prediction. *Bioinformatics*, 36(8): 2547–2553.
- Chen, Y.; Ma, T.; Yang, X.; Wang, J.; Song, B.; and Zeng, X. 2021. MUFFIN: multi-scale feature fusion for drug–drug interaction prediction. *Bioinformatics*, 37(17): 2651–2658.
- Chu, X.; Lin, Y.; Wang, Y.; Wang, L.; Wang, J.; and Gao, J. 2019. MLRDA: A multi-task semi-supervised learning framework for drug-drug interaction prediction. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 4518–4524.
- Cui, L.; Seo, H.; Tabar, M.; Ma, F.; Wang, S.; and Lee, D. 2020. Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 492–502.
- Deepika, S.; and Geetha, T. 2018. A meta-learning framework using representation learning to predict drug–drug interaction. *Journal of biomedical informatics*, 84: 136–147.
- Deng, Y.; Xu, X.; Qiu, Y.; Xia, J.; Zhang, W.; and Liu, S. 2020. A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics*, 36(15): 4316–4322.
- He, C.; Liu, Y.; Li, H.; Zhang, H.; Mao, Y.; Qin, X.; Liu, L.; and Zhang, X. 2022. Multi-type feature fusion based on graph neural network for drug–drug interaction prediction. *BMC bioinformatics*, 23(1): 224.
- Hong, Y.; Luo, P.; Jin, S.; and Liu, X. 2022. LaGAT: link-aware graph attention network for drug–drug interaction prediction. *Bioinformatics*, 38(24): 5406–5412.
- Huang, K.; Xiao, C.; Hoang, T.; Glass, L.; and Sun, J. 2020. Caster: Predicting drug interactions with chemical substructure representation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 702–709.
- Kim, E.; and Nam, H. 2022. DeSIDE-DDI: interpretable prediction of drug–drug interactions using drug-induced gene expressions. *Journal of cheminformatics*, 14(1): 1–12.
- Kipf, T. N.; and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Lee, G.; Park, C.; and Ahn, J. 2019. Novel deep learning model for more accurate prediction of drug–drug interaction effects. *BMC bioinformatics*, 20: 1–8.
- Li, Y.; Hsieh, C.-Y.; Lu, R.; Gong, X.; Wang, X.; Li, P.; Liu, S.; Tian, Y.; Jiang, D.; Yan, J.; et al. 2022. An adaptive graph learning method for automated molecular interactions and properties predictions. *Nature Machine Intelligence*, 4(7): 645–651.
- Lin, S.; Chen, W.; Chen, G.; Zhou, S.; Wei, D.-Q.; and Xiong, Y. 2022a. MDDI-SCL: predicting multi-type drug–drug interactions via supervised contrastive learning. *Journal of Cheminformatics*, 14(1): 1–12.
- Lin, S.; Wang, Y.; Zhang, L.; Chu, Y.; Liu, Y.; Fang, Y.; Jiang, M.; Wang, Q.; Zhao, B.; Xiong, Y.; et al. 2022b. MDF-SA-DDI: predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Briefings in Bioinformatics*, 23(1): bbab421.
- Lin, X.; Quan, Z.; Wang, Z.-J.; Ma, T.; and Zeng, X. 2020. KGNN: Knowledge Graph Neural Network for Drug–Drug Interaction Prediction. In *IJCAI*, volume 380, 2739–2745.
- Lyu, T.; Gao, J.; Tian, L.; Li, Z.; Zhang, P.; and Zhang, J. 2021. MDNN: A Multimodal Deep Neural Network for Predicting Drug–Drug Interaction Events. In *IJCAI*, 3536–3542.
- Ribeiro, L. F.; Saverese, P. H.; and Figueiredo, D. R. 2017. struc2vec: Learning node representations from structural identity. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 385–394.
- Ryu, J. Y.; Kim, H. U.; and Lee, S. Y. 2018. Deep learning improves prediction of drug–drug and drug–food interactions. *Proceedings of the national academy of sciences*, 115(18): E4304–E4311.
- Shi, J.-Y.; Mao, K.-T.; Yu, H.; and Yiu, S.-M. 2019. Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi-nonnegative matrix factorization. *Journal of cheminformatics*, 11(1): 1–16.
- Su, X.; Hu, L.; You, Z.; Hu, P.; and Zhao, B. 2022. Attention-based knowledge graph representation learning for predicting drug–drug interactions. *Briefings in bioinformatics*, 23(3): bbac140.
- Tang, J.; Qu, M.; Wang, M.; Zhang, M.; Yan, J.; and Mei, Q. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on world wide web*, 1067–1077.
- Vilar, S.; Uriarte, E.; Santana, L.; Lorberbaum, T.; Hripcak, G.; Friedman, C.; and Tatonetti, N. P. 2014. Similarity-based modeling in large-scale prediction of drug–drug interactions. *Nature protocols*, 9(9): 2147–2163.
- Wang, D.; Cui, P.; and Zhu, W. 2016. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 1225–1234.
- Wang, H.; Zhao, M.; Xie, X.; Li, W.; and Guo, M. 2019. Knowledge graph convolutional networks for recommender systems. In *The world wide web conference*, 3307–3313.
- Wang, Y.; Wang, J.; Cao, Z.; and Barati Farimani, A. 2022. Molecular contrastive learning of representations via graph

neural networks. *Nature Machine Intelligence*, 4(3): 279–287.

Yu, Y.; Huang, K.; Zhang, C.; Glass, L. M.; Sun, J.; and Xiao, C. 2021. SumGNN: multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics*, 37(18): 2988–2995.

Zhang, X.; Wang, G.; Meng, X.; Wang, S.; Zhang, Y.; Rodriguez-Paton, A.; Wang, J.; and Wang, X. 2022. Molormer: a lightweight self-attention-based method focused on spatial structure of molecular graph for drug–drug interactions prediction. *Briefings in Bioinformatics*, 23(5): bbac296.

Zitnik, M.; Agrawal, M.; and Leskovec, J. 2018. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13): i457–i466.